MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 8: 235–257 (2024)
DOI: 10.3897/mbmg.8.122375

Research Article

Autonomous samplers and environmental DNA metabarcoding: sampling day and primer choice have greatest impact on fish detection probabilities

Devin N. Jones[1], Ben C. Augustine[1], Patrick Hutchins[1], James M. Birch[2], Kevan M. Yamahara[2], Scott Jensen[2], Rodney Richardson[3,4], Regina Trott[3], James R. Campbell[5], Elliott Barnhart[6], Adam J. Sepulveda[1]

1 *U.S. Geological Survey, Northern Rocky Mountain Science Center, Bozeman, MT, USA*
2 *Monterey Bay Aquarium Research Institute, Moss Landing, CA, USA*
3 *University of Maryland, Center for Environmental Science, Appalachian Laboratory, Frostburg, MD, USA*
4 *Western EcoSystems Technology, Inc., Indianapolis, IN, USA*
5 *U.S. Geological Survey Wyoming-Montana Water Science Center, Riverton, WY, USA*
6 *U.S. Geological Survey Wyoming-Montana Water Science Center, Helena, MT, USA*

Corresponding author: Devin N. Jones (dnjones@usgs.gov)

OPEN ACCESS

Academic editor: Bernd Hänfling
Received: 5 March 2024
Accepted: 22 August 2024
Published: 13 September 2024

Citation: Jones DN, Augustine BC, Hutchins P, Birch JM, Yamahara KM, Jensen S, Richardson R, Trott R, Campbell JR, Barnhart E, Sepulveda AJ (2024) Autonomous samplers and environmental DNA metabarcoding: sampling day and primer choice have greatest impact on fish detection probabilities. Metabarcoding and Metagenomics 8: e122375. https://doi.org/10.3897/mbmg.8.122375

Abstract

Unprecedented rates of biodiversity loss and ecosystem function necessitate the use of rapid, efficacious, and cost-effective biomonitoring tools. The combination of autonomous samplers and high throughput sequencing (i.e., "metabarcoding") of environmental DNA (eDNA) samples enables characterization of entire communities at high frequency and can be an important tool for conservation and management, allowing researchers to track fluctuations in biodiversity. We deployed two autonomous samplers at two U.S. Geological Survey streamgage sites in the upper Snake River (Wyoming and Idaho, USA) to collect eDNA samples from July-September 2021 and 2022 to characterize fish diversity. We used a probabilistic approach to evaluate the effects of water temperature, water discharge, filter pore size, water volume filtered, number of samples collected, timing, and primers on the probability of detecting eDNA from fish species known to be present. We detected eDNA from 13/15 species present in these areas of the Snake River. Overall, we did not find evidence that filter pore size, water volume filtered, water discharge, and water temperature affected the probability of detecting fish species' eDNA. By contrast, primers and sampling day affected fish detection probabilities, indicating that primer choice and sampling day can either over- or under- estimate species diversity. These results indicate that users would ideally consider sampling on non-consecutive days and which primer set will maximize species detections.

Key words: Autonomous samplers, biomonitoring, eDNA, metabarcoding, USGS streamgage

Introduction

Biodiversity monitoring informs species conservation, ecosystem service management, and progress toward major national and international policy initiatives (e.g., Díaz et al. 2019; Aichi Biodiversity Targets 2020). However, there is a mismatch between the scale of biomonitoring needed in our rapidly changing world and the biomonitoring tools currently available. Conventional biomonitoring tools (e.g., nets and traps) are limited to short temporal windows, low sampling frequencies, and can be restricted to a subset of species present in the environment (Moussy et al. 2022). These shortcomings can result in a misrepresentation of biodiversity and inefficient allocation of limited resources to achieve conservation outcomes. There is a need for the development and implementation of additional tools that overcome these limitations.

Environmental DNA (eDNA) metabarcoding analysis is a biomonitoring approach that allows for biological communities to be described from short sections of DNA that occur in easy-to-collect environmental samples. The approach is accurate, rapid, cost-effective, non-invasive and can be consistently and broadly applied (Ruppert et al. 2019). Nevertheless, eDNA metabarcoding, like any other tool, provides imperfect representation of biological life present due to constraints imposed by a myriad of field and laboratory factors. Here, we focus on the implementation of new technology and sample design considerations to help overcome one of these factors that has been generally neglected in previous studies—temporal heterogeneity of eDNA in the environment (e.g., Mathieu et al. 2020; Hallam et al. 2023).

The amount of eDNA available to be sampled at a site is temporally variable because it is a product of dynamic biological and environmental interactions. Organisms release complex mixtures of DNA in the environment at inconsistent rates and amounts that are associated with taxonomy, size, activity, and phenology among other things (Barnes and Turner 2016). Abiotic factors and microbial activity break down this eDNA into smaller fragments while it is being adsorbed and transported in multiple dimensions (Jo et al. 2020). Therefore, an eDNA sample is often a snapshot of the subset of eDNA present at a specific time and space and is likely to provide incomplete representation of biological communities (Jensen et al. 2022). Multiple samples over time can be required for a more complete biomonitoring picture and to have confidence that eDNA-inferred changes in community composition across time points are indicative of species turnover rather than sampling bias. However, human resource limitations pose challenges to the increased temporal sampling needed to complete this picture.

Autonomous eDNA sampling instrumentation presents an opportunity to overcome human resource limitations for ecosystem monitoring over time and space. Autonomous eDNA instruments, like Monterey Bay Aquarium Research Institute's environmental sample processor (ESP) platforms (Moss Landing, CA, USA), can collect and preserve dozens of eDNA water samples without a human operator (e.g., Scholin et al. 2017; Sepulveda et al. 2021; Preston et al. 2023). Thus, eDNA data can be acquired at sites and times that are less amenable to manual sampling, such as remote sites, nights, and hazardous environmental conditions. This technology has great promise to reduce biomonitoring misrepresentation associated with the temporal heterogeneity of eDNA in the

environment because samples can be collected at finer temporal frequencies (e.g., sub-daily; Searcy et al. 2022) for longer durations (e.g., weeks to months) than most manual eDNA sampling efforts. However, continuous improvement of study design is needed to ensure temporal sampling frequency is appropriate for the system because sample number is still limited, and autonomous sampler deployment and analysis are costly.

Understanding about appropriate eDNA temporal sampling frequencies for biomonitoring is minimal relative to many eDNA workflow methods (Mathieu et al. 2020), likely because eDNA autonomous sampling is a newer tool and meaningful insight on this topic is largely limited to field studies rather than laboratory experiments. The number of temporal studies is increasing, but is still largely limited to longer (i.e., months to seasons) rather than shorter time scales (i.e., hours to weeks) (Bista et al. 2017; Beentjes et al. 2019; Handley et al. 2019; Hayami et al. 2020; Mariani et al. 2021; Sales et al. 2021; Hervé et al. 2022). Sampling at longer time scales can capture eDNA from taxa that are seasonally present (e.g., migration) and that have increased activity associated with phenology (e.g., reproduction). By contrast, sampling at shorter time scales is more likely to capture eDNA from low density species that quickly degrades and whose activity and habitat use vary diurnally. For example, Jensen et al. (2022) showed that fish species eDNA detections were highly dynamic throughout a 32-hour period, with eDNA-inferred species richness varying substantially at an hourly rate with some species having a greater number of detections mid-day while other species had a greater number of detections at night. Thus, sampling at shorter time scales that more accurately represents community composition at a site may be critical to increasing confidence in species turnover inferred from eDNA samples collected at longer time scales, which is a primary objective of many biomonitoring programs (Bálint et al. 2018). As resources limit the amount of sampling that can be done for biomonitoring, tradeoffs between short- and long-term sampling periods should be considered.

To gain insight on temporal sampling tradeoffs for eDNA biomonitoring programs, we leveraged existing eDNA samples that were collected for invasive species surveillance by autonomous samplers at sub-daily frequencies for three months from summer (July) to autumn (September) seasons in 2021 and 2022. The autonomous samplers were deployed at U.S. Geological Survey (USGS) streamgages immediately downstream of two reservoirs in the Snake River (Idaho and Wyoming, USA). These waters contain fish species known to exhibit variation in daily and seasonal activities, such as cutthroat trout (*Oncorhychus clarkii*) that reproduce in the summer and lake trout (*Salvelinus namaycush*) that reproduce in the fall (Gunn 1995; Henderson et al. 2000). We used eDNA metabarcoding approaches on these existing samples to describe fish community composition in the upstream reservoirs. We predicted that 1) fish eDNA detections will vary sub-daily, daily, and seasonally and 2) an increased number of samples collected per day will increase the number of fish species eDNA detections. We also evaluated how temporal sampling tradeoffs could be influenced by other study design decisions like filter pore size and primer choice while accounting for covariates like water temperature and water volume filtered. Our analyses highlight the importance of sampling design choices in the field and lab when using eDNA samples for biodiversity monitoring.

Methods

Site descriptions and sample collection

We leveraged existing eDNA samples associated with a separate project that deployed Monterey Bay Aquarium Research Institute's second-generation ESPs for targeted eDNA surveillance of aquatic invasive species in the upper Snake River (Wyoming and Idaho, USA). While our sample set and analyses are limited by the study design of this separate project, the high frequency of sample collection provides a unique opportunity to address knowledge gaps in eDNA metabarcoding study design. Second-generation ESPs (as described in Sepulveda et al. 2020) were deployed on July 12 through September 22, 2021. One ESP was located at USGS streamgage 13011000 (Snake River near Moran, WY; 43°51'31"N, 110°35'09"W; U.S. Geological Survey 2023) on the Snake River near the outflow of Jackson Lake (ESP location hereafter referred to as "Jackson Lake") and the other ESP was located at USGS streamgage 13032500 (Snake River at Irwin, ID; 43°21'03"N, 111°13'08"W; U.S. Geological Survey 2023) on the Snake River near the outflow of Palisades Reservoir (ESP location hereafter referred to as "Palisades Reservoir") located in Wyoming and Idaho, USA, respectively (Fig. 1). USGS streamgage 13011000 is located approximately 330 m downstream of Jackson Lake Dam and USGS streamgage 13032500 is located approximately 2.2 km downstream of the Palisades Dam. The following year, the same two ESPs were deployed at the same locations on July 12 through September 10, 2022.

We used these eDNA samples collected by ESPs at streamgage sites to infer fish community composition at the two upstream reservoirs. Previous research found that streamgage sites on large rivers, including those in this study, are adequately located for biosurveillance of upstream reservoirs (Sepulveda et al. 2019, 2021) as eDNA water samples can integrate biological information over large spatial scales (e.g., Pont et al. 2018). Fifteen fish species (Table 1) have been documented across the two upstream reservoirs by traditional capture methods. All 15 fish species are present across both sites. Both streamgage sites experience similar patterns in water discharge and temperature, with discharge peaking in the early summer from snowmelt runoff, and temperatures increasing throughout July and beginning to cool in September (Suppl. material 1: table S1, figs S1, S2).

Water samples were filtered at daily or sub-daily time increments (Suppl. material 1: table S2). Negative controls consisting of molecular-grade water were collected before deployment and after retrieval for each ESP. Once the target volume (2L) was filtered or the filter was clogged with biomass, filtering stopped, excess water was cleared, and filters were preserved with RNAlater (ThermoFisher Scientific, Waltham, MA, USA). Preserved filters were stored at the ESP temperature, which were similar to ambient air temperatures. The ESPs were housed inside a protective case within USGS streamgages and therefore filters were stored out of direct sunlight. Because the ESPs recorded the volume of water filtered for each sample, we were able to test the effect of water volume filtered (ranging from 199.8 mL to 2,025 mL with an average of 774.3 mL) on the probability of species detections.

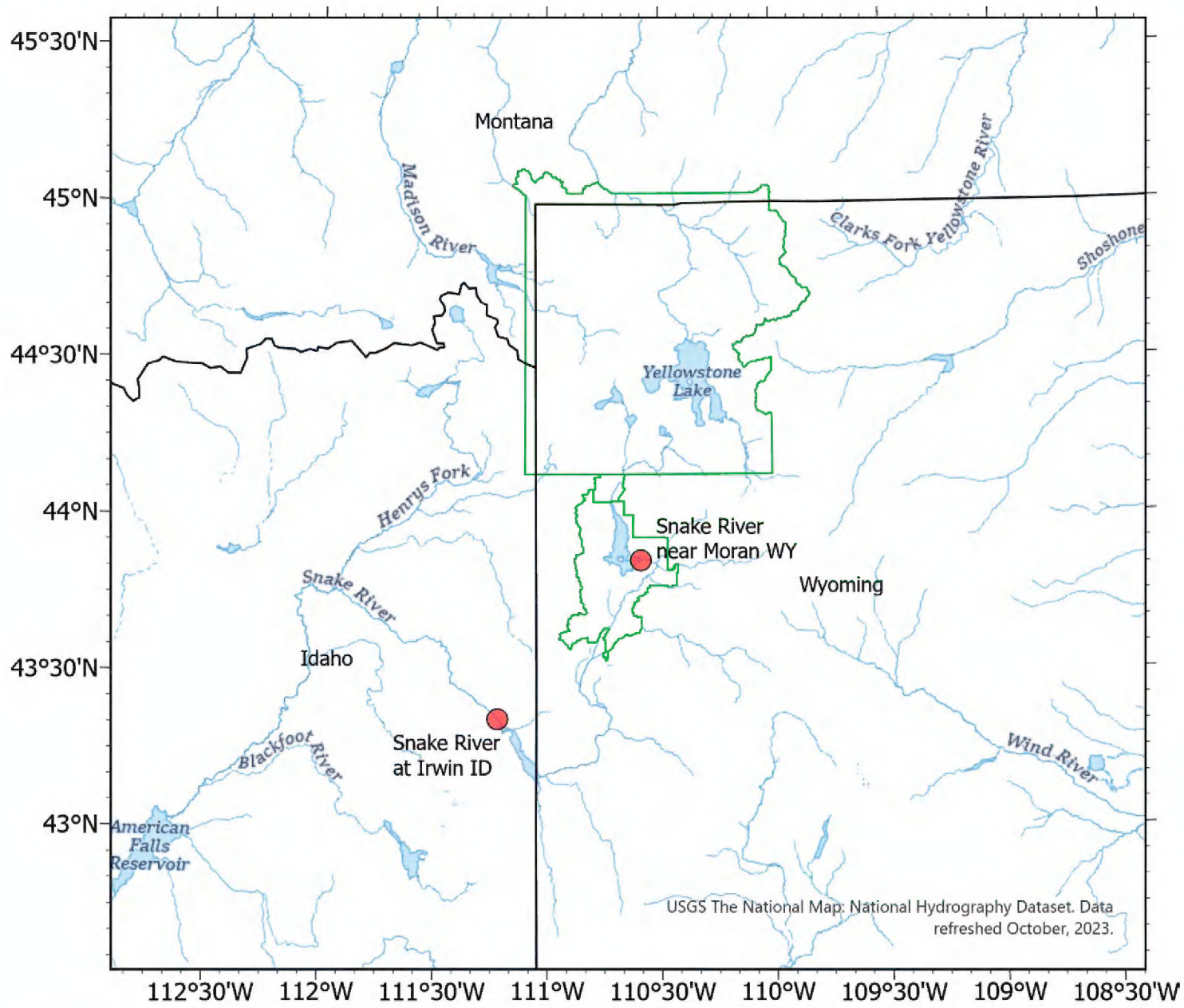

Figure 1. Map of the Upper Snake River in Wyoming and Idaho, USA. ESP locations are shown with pink circles. Green lines denote national parks (Yellowstone National Park and Grand Tetons National Park). Black lines indicate state borders.

Table 1. Fish species detected by each primer set for 2021 and 2022 samples. An (M) next to the species name indicates that genetic material from this species was included in our mock community samples. Under each primer column (with the two sampling years included), a JL indicated that a species' DNA was detected at the Jackson Lake site and a PR indicates that a species' DNA was detected at the Palisades Reservoir site. The letter M indicates a species' DNA was detected in the mock community samples.

Species Name	Common Name	Primers Used (Year of Sample Collection)		
		MiFish (2021)	MiFish (2022)	AcMDB (2022)
Catostomus ardens (M)	Utah sucker			JL, PR, M
Catostomus discobolus	Bluehead sucker	JL, PR		PR
Catostomus platyrhynchus	Mountain sucker			
Cottus bairdii (M)	Mottled sculpin	JL, PR, M	JL, PR, M	M
Cottus beldingii	Paiute sculpin			
Gila atraria (M)	Utah chub		JL, PR, M	
Oncorhynchus clarkii	Cutthroat trout	JL, PR	JL, PR	JL, PR
Oncorhynchus mykiss (M)	Rainbow trout	JL, PR, M	JL, PR, M	JL, PR, M
Oncorhynchus nerka (M)	Kokanee	JL, PR	JL, PR, M	JL, M
Prosopium williamsoni (M)	Mountain whitefish	JL, PR, M	JL, PR, M	JL, PR, M
Rhinichthys cataractae	Longnose dace	JL	JL, PR	PR
Rhinichthys osculus (M)	Speckled dace	JL, PR	PR, M	JL, PR, M
Richardsonius balteatus (M)	Redside shiner	JL, PR, M	JL, PR, M	JL, PR, M
Salmo trutta (M)	Brown trout	JL, PR	JL, PR, M	JL, PR, M
Salvelinus namaycush (M)	Lake trout	JL, PR, M	JL, PR, M	M

We used available data from the U.S. Geological Survey's National Water Information System to describe the discharge (daily mean, m^3/sec; U.S. Geological Survey 2023) and water temperature (daily mean, °C) associated with each streamgage site. We supplemented these data with additional hourly water temperature data by deploying Onset HOBO™ Pendant (Onset, Bourne, MA, USA) temperature loggers at the point of ESP water intake, which recorded temperatures every hour. We used water daily mean discharge rates collected by USGS streamgages. We used these data to assess the effect of these two environmental factors on probability of eDNA detection for each fish species.

Sample selection and sample subsets

Filter pore size

We conducted an initial assessment of the effects of filter pore size on fish eDNA detections by using a subset of ESP samples from 2021 that were collected using 0.45, 1.2, and 5 µm mixed cellulose ester (MCE) Millipore filters (Sigma-Aldrich, Darmstadt, Germany). Samples for filter pore size comparison were collected and filtered consecutively as the ESPs cannot collect multiple samples simultaneously. The ESP collected this subset of samples in order of size 0.45 µm, 1.2 µm, then 5 µm repeatedly at 3-hr intervals across four days in September 2021 for a total of 18 samples (three replicates of each filter pore size per site). Samples collected for filter pore size comparison were collected on September 10–12, 2021, at the Jackson Lake site and on September 12–14, 2021, at the Palisades Reservoir site. We did not use samples from 2022 to assess filter pore size on fish eDNA detections as we used a different filter type of one pore size (see 'Sample frequency and timing' below).

Primer choice

For samples collected in 2021, we amplified and sequenced DNA using the MiFish-U primer set (hereafter referred to as 'MiFish'; Miya et al. 2015) given the well-documented performance of MiFish to describe freshwater fish species richness (Miya et al. 2020). Because we were unable to detect eDNA from several known fish species using the MiFish primer set on the 2021 samples, we chose to use an additional 12S primer set: AcMDB-07 (hereafter referred to as 'AcMDB'; Bylemans et al. 2018) to compare fish eDNA detections for samples collected in 2022. Primer sequences used are provided in the Suppl. material 1: table S3. Zhang et al. (2020) showed that the AcMDB primer set was able to detect the greatest number of fish species compared to other primer sets (including MiFish, which also performed well). Samples selected for extraction were amplified by both sets of primers in parallel for paired comparison. These paired samples from 2022 were the only samples used to answer questions about differences in results between primer sets.

Sample frequency and timing

To assess the optimal number of samples needed to obtain reliable estimates of detection and the impact of sub-daily, daily, and seasonal timing on expected detections, we used both the 2021 and 2022 datasets. However,

these datasets were analyzed separately because of methodological differences associated with the study design of the initial project; 2021 samples used MCE filters of variable pore size whereas 2022 samples used polyethersulfone (PES) filters (Sterlitech, Kent, WA, USA) of the same pore size (1.2 µm only).

DNA extraction and library preparation

Upon retrieval from the ESPs in mid-September, filters were placed into Qiagen Investigator Lyse and Spin Baskets (Qiagen, Germany) containing 180 µL of Qiagen buffer ATL and 20 µL of proteinase K. Samples were then transported back to the U.S. Geological Survey Northern Rocky Mountain Science Center (NOROCK) in Bozeman, Montana, where they were incubated at 56 °C for twelve hours. DNA extraction was carried out with Qiagen DNEasy Blood and Tissue kits according to the manufacturer's instructions. The final elution volume was 100 µL for 2021 samples and 50 µL for 2022 samples. Elution volume was lower for 2022 since DNA concentrations and quality were low for 2021 samples. DNA extracts were stored at -80 °C until further processing.

For samples collected in 2021, each sample was amplified in triplicate with untailed MiFish primers and triplicates were pooled and shipped frozen to the University of Maryland Center for Environmental Science Appalachian Laboratory. Sample DNA extracts were prepared for Illumina sequencing using the methods of Richardson et al. (2019). Details are provided in Suppl. material 1.

Samples collected in 2022 were amplified in triplicate with two primer sets (MiFish and AcMDB) in simplex, after which each replicate amplicon was amplified using respective tailed versions of those same primers that were tagged on both the forward and reverse primers (sequences given in Suppl. material 1: table S2). Sample replicate amplicons of the same primer set were then pooled, cleaned using magnetic beads, and eluted in Tris-HCL pH 8.5. DNA concentration of each elution was measured on a Qubit fluorometer (ThermoFisher Scientific, Waltham, MA, USA) and samples were normalized such that the final DNA concentration of each elution was between 2–5 ng μL^{-1}. Cleaned and normalized samples were then sent to the Michigan State University Research Technology and Support Facility for indexing and sequencing using Illumina MiSeq (San Diego, CA, USA). Details for sequencing preparation of 2022 samples are provided in the Suppl. material 1.

In addition to the aforementioned negative field controls, each 96-well PCR plate contained the following controls, which were carried through tailed and untailed amplifications and sequencing: a no-template control using 1 µL of sterile water per reaction, 1 µL of pooled extraction blanks (one extraction blank for every batch of 47 field samples and this was only included in sequencing for the 2022 samples), and one (for 2021 samples) or three (2022 samples) replicate wells per plate of a synthetic mock community comprised of $1e^3$ 12s gene copies of each of ten fish species (see Suppl. material 2 for species and FASTA sequences used to make synthetic DNA). Synthetic DNA for the mock community was synthesized as gblocks by Integrated DNA Technologies (www.idtdna.com).

Bioinformatic analysis of sequences

To generate updated reference sequence databases for each genetic marker, we downloaded all available 12S sequence data from NCBI on March 27th, 2023, and used MetaCurator (v1.0.1; Richardson et al. 2020) to identify the MiFish and AcMDB amplicon regions of interest, trim away extraneous 3' and 5' sequence data and perform taxonomy-aware de-replication of the resulting references. MAFFT v7.270 (Katoh et al. 2002), HMMER3 v3.1b2 (Eddy 2011) and VSEARCH v2.8.1 were used as dependencies. The final curated databases contained representative sequences for 18,941 species from 6,746 genera for MiFish and 37,216 species from 13,855 genera for AcMDB.

We aligned eDNA sequence data against these reference databases using VSEARCH v2.8.1 (Rognes et al. 2016). For each marker, paired ends were merged, priming sites were clipped from the 3' and 5' ends and sequences less than 100 bp in length were discarded. Semi-global top-hit alignment was performed with a minimum query coverage requirement of 80%. Sequences with a percent identity match of ≥ 98% were annotated to species according to the taxonomic identity of the closest reference. An example of bioinformatic workflow can be found in the code available in Crone et al. (2023).

Decontaminants were identified and proportionally removed with the R package *microDecon* (McKnight et al. 2019). Reads from each set of samples (as defined by site, primer, and year of collection) were decontaminated separately with their corresponding negative controls. All reads that were removed from samples can be found in Suppl. material 3. For the final dataset, we only retained operational taxonomic units (OTUs) identified as species that are known to occur at these sites (i.e., fish species that have been detected with traditional capture methods) because our objective was to compare observed fish species detections to expected fish species.

Model description and parameter estimation

We chose to model the data at the species-by-sample level, instead of the number of species detected per sample, which may be useful for eDNA metabarcoding studies moving forward. Our approach is similar to that of Dorazio et al. (2006), which uses an occupancy model to produce site-level species richness estimates, accounting for imperfect detection. However, given the possibility that the detection probability below the lakes may be zero for some species occupying the lake due to inappropriate primers for those species, or removal of the eDNA before it reaches the sampling location, we modeled detection and detected species richness. Though occupancy and detection are confounded, we rely on prior knowledge of the species communities in these lakes to assume all species we targeted occupy the reservoirs. This type of analysis allows for species-specific detection probabilities and the incorporation of covariates, interactions with covariates, and random effects like site, season, water volume filtered, and others. We modeled our data using mixed-effects logistic regression (Gelman and Hill 2006) with separate models for 2021 and 2022. Because the ESPs were unable to collect true replicate samples at the sub-daily level, we were unable to assess the effect of sample collection time on species detection probabilities. However, we were able to use the sub-daily

samples as day-level replicates. Year effects were not estimable due to different filter types used in each year and because only 2022 used more than one primer. For 2021, our linear predictor was

$$\text{logit}(p_i) = \beta_{0species_i} + \beta_1 \times volume_i + \beta_{2\,filter_i} + \eta_{species_i site_i month_i} + \zeta_{day_i} \quad (1)$$

where $\beta_{0species_i}$ is the intercept for the species of observation i, $volume_i$ is the sample volume of observation i, and $filter_i$ is the filter pore size of observation i, with filter sizes 1.2 µm and 5 µm treated as offsets from 0.45 µm (i.e., $\beta_{21} = 0$). Next, $\eta_{species_i site_i month_i}$ is the species by site by month effect of observation i and ζ_{day_i} is the day effect of observation i. We modeled both species by site by month effects and day effects with non-centered random intercepts (Betancourt and Girolami 2015). We assumed $\eta_{i,j,k} = \varphi\sigma\eta$ for species i at site j in month k where φ ~ Normal(0,1) and ση is the species by site by month standard deviation. Then for day effects, we assumed $\zeta_d = \psi_d\sigma\zeta$ for day d where $\psi_d \sim$ Normal(0, 1) and σζ is the standard deviation for day.

For 2022, our linear predictor was

$$\text{logit}(p_i) = \beta_{0species_i primer_i} + \beta_1 \times volume_i + \eta_{species_i site_i month_i} + \zeta_{day_i} \quad (2)$$

where $\beta_{0species_i primer_i}$ is the intercept for the species by primer combination of observation i and all other terms are the same as in the 2021 model.

For both models, we derived the expected number of species detected per sample, assuming independence across species, as a function of all factor levels in the model for each year. In 2021, we derived the cumulative detection probability for species s at site j in month k and filter pore size f for n = 1, …, 10 hypothetical samples following

$$p(n)_{s,j,k,f} = 1 - (1 - p_{s,j,k,f})^n \quad (3)$$

where $p_{s,j,k,f}$ is the detection probability of species s at site j in month k for filter size f. Then, we computed the expected number of species detected as a function of n samples for these factor levels following

$$\lambda(n)_{j,k,f} = \sum_s^{Nspecies} p(n)_{s,j,k,f} \quad (4)$$

In 2022, we derived the expected number of species detected per sample in the same manner, except these computations were stratified by primer, site, and month instead of filter, site, and month. Again, this approach is analogous to that of Dorazio et al. (2006) except we compute detected, rather than true species richness.

We fit both models via Markov chain Monte Carlo (MCMC) using the NIMBLE software (de Valpine et al. 2017) in R (version 4.3.1). We used uninformative priors for all model parameters, using Logistic(0,1) priors for intercepts on the logit scale and Normal(0,10) for fixed coefficients. We used Uniform(0,10) priors for both random effect standard deviation parameters. For each model, we ran 3 chains of 50,000 iterations, thinned by 5, with 1,000 thinned samples discarded as burn-in. Convergence was then assessed using the Gelman-Rubin statistic (often referred to as $\hat{R}$; Gelman and Rubin 1992) and verified that the

95% confidence interval upper bound was less than 1.1. We computed posterior means for point estimates and 95% Highest Posterior Density (HPD) intervals for interval estimates (Gelman and Hill 2006). We explored the effects of water volume filtered, water discharge, and temperature and found that temperature and discharge estimates were unstable due to correlation with each other (and month) and were therefore omitted. We included water volume filtered in the model even though water volume estimates were centered around zero for 2021 data and moderately positive in 2022, but still overlapping zero. In this study, we were only able to test the effect of water volume filtered for the limited ranges of volumes collected by the ESP.

Results

We sequenced a total of 176 field samples with the number of samples collected per day ranging from one to six, eight mock community samples, and 20 negative controls (including field, extraction, PCR, and sequencing blanks). In 2021, we sequenced 45 samples from Jackson Lake and 45 samples from Palisades Reservoir collected over 12 days. In 2022, we sequenced 46 samples from Jackson Lake (n_{MiFish} = 46, n_{ACMDB} = 46) and 40 samples from Palisades Reservoir (n_{MiFish} = 40, n_{ACMDB} = 40) collected over 16 days (breakdown of samples collected at each site per day listed in Suppl. material 1: table S3). Across all datasets, five classes of eukaryotes were detected (Actinopteri, Aves, Mammalia, Bacillariophyceae, and Insecta) with a total of 8,058,780 reads. We retained only fish (Actinopteri) detections, resulting in 63 unique species for the 2021 dataset (2,041,818 reads), 71 unique species for the 2022 MiFish dataset (3,837,411 reads), and 69 unique species for the 2022 AcMDB dataset (2,400,946 reads). Finally, we filtered the dataset further to include only species known to be in the Snake River, resulting in a total of 3,709,008 reads across 11 detected species (Table 1) for both sites and years in our final dataset.

Differences among filter pore sizes and optimized sampling

We found overlapping probabilities of detecting eDNA from the same number of fish species for each of the filter pore sizes, indicating no effect of filter pore size on results (Fig. 2). The number of samples needed to detect eDNA of most fish species present reached an asymptote beginning at 2–3 samples per day (Fig. 2). These results are consistent between sites and months.

Differences between primer sets

The two different primers were able to detect eDNA from a total of 13 out of 15 fish species known to occur (Table 1). Neither primer set was able to detect *Catostomus platyrhynchus* or *Cottus beldingii*. For one (*C. platyrhynchus*) of the two undetected fish species, which had no 12S sequence data available, as well as *C. ardens*, we were able to find specimens (i.e., fin clips from fish in the upper Snake River) to sequence the entire mitochondria. These mitochondrial sequences can be found under the GenBank submission number BankIt2662303 (#s PP116420–PP116425).

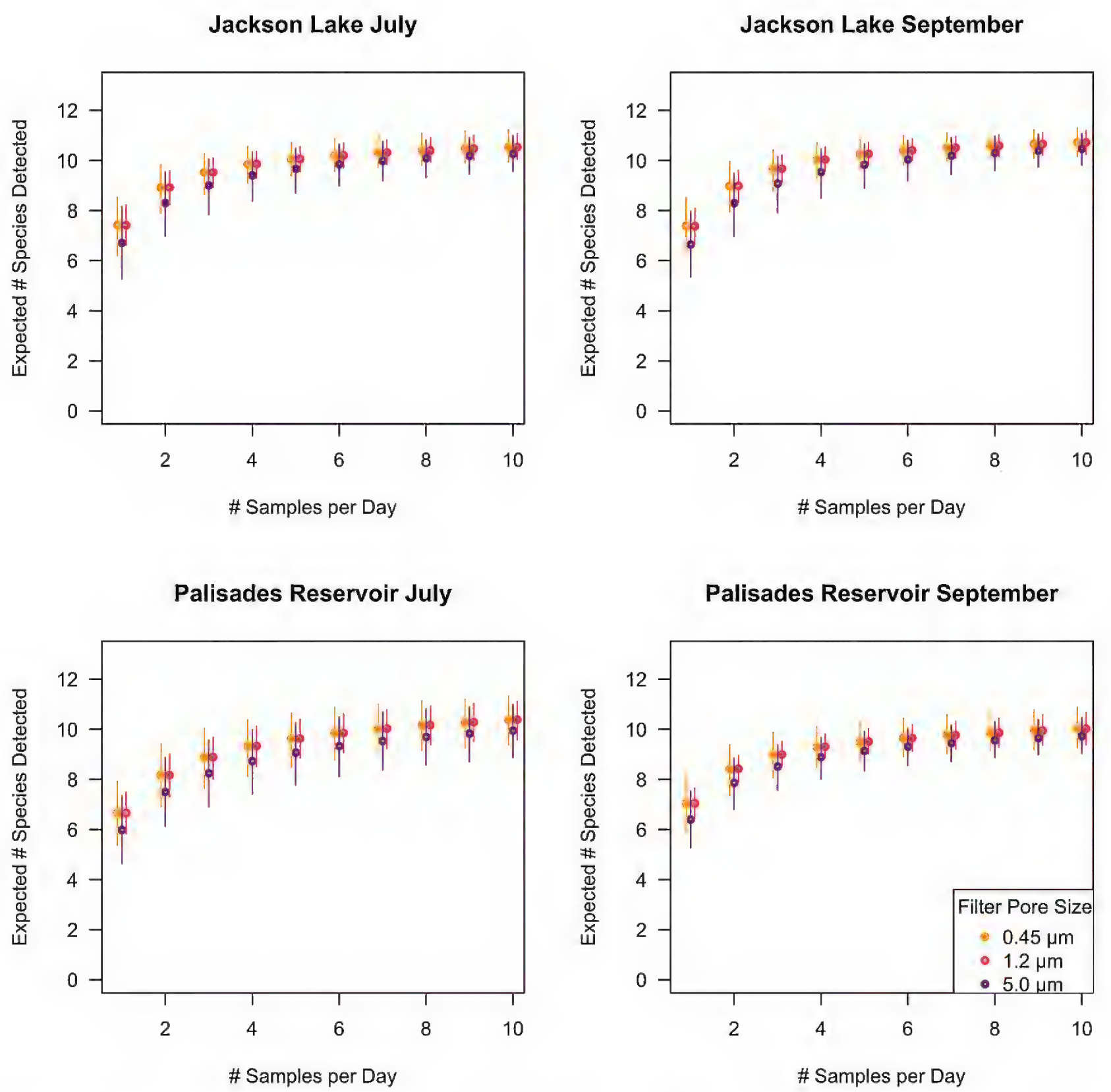

Figure 2. Detected species accumulation curves showing the point and interval estimates for the expected number of species detected from eDNA based on the number of samples collected for each site during each month in 2021. Colors indicate the three different filter pore sizes tested. Bars represent 95% confidence intervals.

The MiFish primer set was able to detect eDNA from three species (*C. bairdii*, *G. atraria*, and *Salvelinus namaycush*) that AcMDB was unable to detect. Interestingly, the AcMDB primers were able to detect *C. bairdii* and *S. namaycush* in the mock communities. By contrast, the AcMDB primer set was able to detect one additional species (*C. ardens*) in the field samples that the MiFish primer set did not. We found primer choice resulted in species-specific differences in detection using samples from 2022 to assess the difference in fish detection probabilities. Five fish species were detected with both primer sets; six species were detected with only one primer set (Fig. 3, Suppl. material 1: table S5). The AcMDB primer set had a greater probability of detecting eDNA from *C. ardens* and *R. osculus*, whereas the MiFish primer set had a greater probability of detecting eDNA from *C. bairdii*, *G. atraria*, *O. nerka*, *R. cataractae*, and *S. namaycush* (Fig. 3). AcMDB primers were able to detect *O. nerka* in a single sample and only at low levels (i.e., 3 reads in that sample).

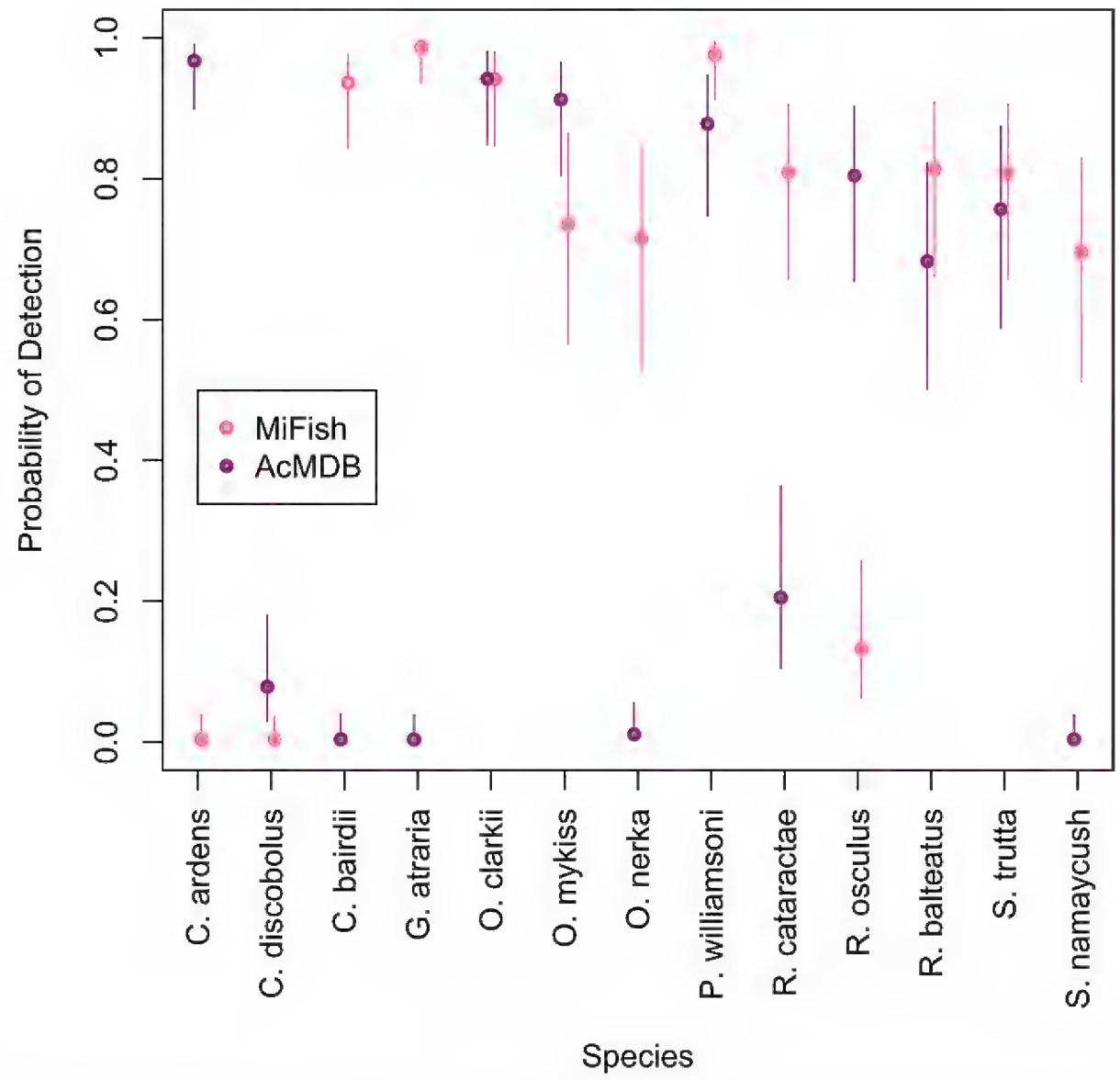

Figure 3. Detection probabilities of each primer for each fish species detected from eDNA in 2022. AcMDB primer detection probabilities are shown in purple for each species, and the MiFish primer detection probabilities are shown in pink for each species. Bars represent 95% confidence intervals.

Effects of timing on sampling

We found that detection probabilities varied by day but lacked a consistent or obvious temporal pattern. Days with 95% credible intervals that do not overlap with zero indicate that the eDNA detection probabilities of all species were higher or lower on those days compared to the average day, implying that either more or fewer number of species' eDNA will be detected on those days (Fig. 4). There was a single day (day 208; July 27, 2022) in July at the Jackson Lake streamgage where fish species' eDNA detections were lower than average. Similarly, there was a single day (day 208) in July at the Palisades Reservoir streamgage with below average fish species detection probabilities. This incidence of below average detection probability at Palisades Reservoir was immediately followed by a day of above average detection probability (day 209; July 28, 2022). There were also two consecutive days in September at the Palisades Reservoir streamgages site when detection probabilities were higher than average (days 249–250; September 6–7, 2022). We did not find a difference in detection probabilities for any fish species between summer and autumn sampling.

Discussion

In light of global declines in biodiversity, it is important to understand how to integrate eDNA sampling into robust monitoring frameworks. Biomonitoring programs are challenged to balance tradeoffs in study design choices, such as sampling frequency and duration since human and funding resources are

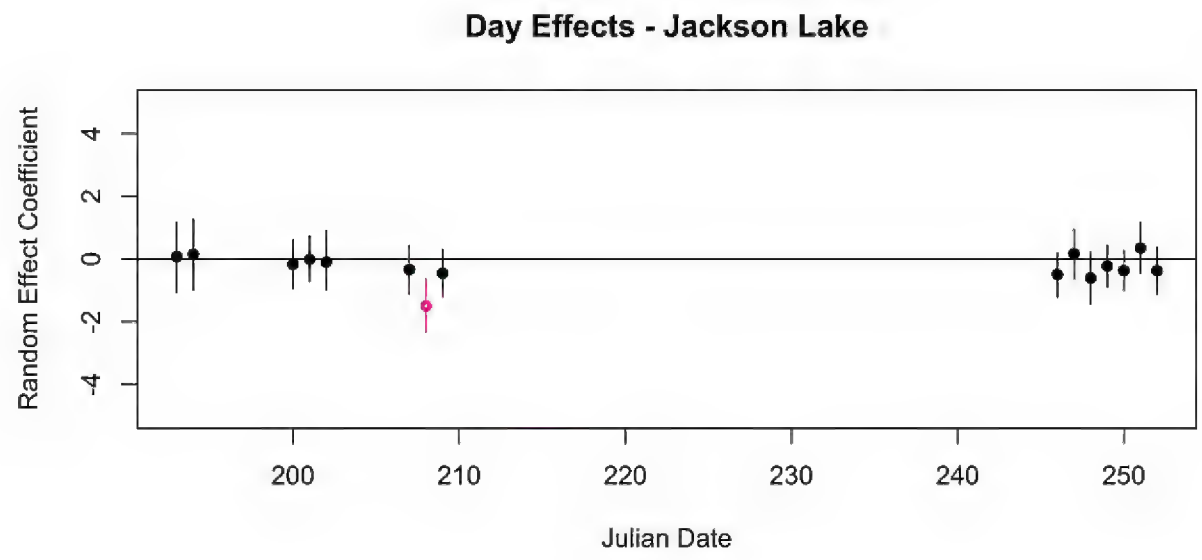

Figure 4. Plots of the detection probability random effect point and interval estimates for each sampling day by site for 2022 samples. Pink points highlight daily detection probability estimate intervals that do not overlap with zero. Bars represent 95% confidence intervals.

limited. Combining eDNA metabarcoding with autonomous samplers into a biomonitoring program provides a new way to reduce these limitations, but study design tradeoffs must still be considered and optimized to minimize the misrepresentation of communities. Here, we considered how sample timing frequency, filter pore size, and primer choice influence inferences of fish community composition. Taken together, results from this study show that, of the factors we considered, primer choice and sampling date had the greatest impact on fish species eDNA detections.

High frequency samples collected autonomously across months from USGS streamgage sites were able to detect eDNA from 13 of 15 fish species known to occur at two upstream reservoirs. Though these data provide an imperfect representation of the known fish assemblage, our results underscore the relatively minimal field effort needed for eDNA-based biomonitoring when using autonomous samplers. Traditional assessments can require several biological technicians employing multiple, potentially destructive methods (e.g., gill nets and electrofishing) at multiple sites and depths and can still result in biased findings. Comparable manual eDNA sampling would have required dozens of site visits. Our assessment required approximately six hours of field time at each of two sites for ESP deployment and retrieval. Technological advancements and standardization are likely to accelerate workflows such that eDNA autonomous platform-based programs can respond to the scale of biomonitoring needed in our rapidly changing world. An outstanding example of this promise is the use of uncrewed surface vessels fitted with autonomous platforms

that were able to collect eDNA samples over a 4200-km, 29-day transit in the northeastern Pacific Ocean (Preston et al. 2023).

Technological advancements like autonomous platforms cannot be used to their full potential in biomonitoring programs if genetic reference libraries are insufficient. Inadequate genetic reference library coverage likely limited our ability to provide a complete picture of Jackson Lake and Palisades Reservoir fish assemblages. Of the two species' eDNA we did not detect, one of these fish species had no 12S sequences available, whereas the other one had only six 12S sequences available. The use of multiple primers did not resolve this issue. The problem of incomplete reference libraries is nearly ubiquitous in eDNA biomonitoring efforts (Schenekar et al. 2020; Keck et al. 2023). Solutions include development of curated, regional databases tied to vouchered specimens (e.g., Dziedzic et al. 2023), expansion of and improvements to public reference databases (e.g., Westfall et al. 2023), or embracing taxonomy-free approaches (e.g., Pawlowski et al. 2021).

Our results emphasize that primer choice is one of the most influential considerations in eDNA metabarcoding study design, as has been demonstrated many times over (e.g., Hajibabaei et al. 2019). Even though both of our primers targeted the 12S region, the MiFish primers were able to detect eDNA from four more species than the AcMDB primers. Given that all but one fish species (*R. osculus*) detected with the MiFish primers had a detection probability greater than 0.6, most species' eDNA should be detected with only a few samples collected per day given that the eDNA is available to be sampled. However, the ability of the AcMDB primers to detect an additional two species in the mock communities compared to field samples indicates that these primers may be more efficient at detecting fish DNA under ideal conditions (e.g., no inhibition and ample template). As eDNA can degrade quickly in the environment, shorter amplicons (i.e., < 200 bp) may be more easily recovered in eDNA samples (Hajibabaei et al. 2006; Meusnier et al. 2008). The MiFish primers target a ~170 bp fragment of the 12S mitochondrial region, whereas the AcMDB primers target a fragment nearly twice the length (~281 bp), which may have hampered our eDNA detection of some species in field samples. There was a benefit to using multiple primers as we were able to detect eDNA from one additional species, but the costs of additional analyses should be considered. Given the impact of primer choice on metabarcoding results, it is important to consider which primer set will optimize species detections to most accurately represent the community of interest.

Implementation of autonomous platforms facilitates high frequency eDNA sampling, which can provide insight into changes in an organism's ecology. Indeed, previous research has shown that sub-daily measurements of eDNA concentrations may reflect changes in daily behavior and activity throughout a photoperiod (e.g., Searcy et al. 2022), and similarly, longer-term high frequency sampling may give insight into seasonal behaviors such as migration or reproduction (e.g., Tillotson et al. 2018; Hallam et al. 2023). Even when accounting for primers, the number of reads for each species was highly variable throughout the day (see example in Suppl. material 1: fig. S3), but because we did not have sample replicates for time of day, we were unable to test the effect of time of day on species detection probabilities. While other eDNA studies have shown that sub-daily sampling mimics diel behavior of fish species (e.g., Jensen et al. 2022; Searcy et al. 2022), our results did not show consistent patterns in the number of reads for any fish species throughout a given day (Suppl. material 1: fig. S3) and in fact,

more closely resemble the sub-daily patchiness in detections seen in Searcy et al. (2022). Therefore, our results did not indicate diel differences in detection, even though several of these fish species are known to have diel variation in activity (e.g., Nowak and Quinn 2002). For example, some species of suckers and dace tend to exhibit more diurnal behavior, whereas some chub species have been shown to be crepuscular (Reebs et al. 1995). Similarly, rainbow trout are likely more diurnal in contrast to cutthroat trout (Nowak and Quinn 2002). Generally, a fish species' diel behavior may vary depending on factors like temperature, age, and food availability. Lack of diel variation could be true or could be a function of sampling downstream of bottom-releasing reservoir dams, where hydraulics may mix eDNA and decouple ecological relationships between eDNA signal and timing at short temporal scales. Hallam et al. (2023) also found low dissimilarity in inferred fish community composition at sub-daily scales and suggested that eDNA mixing caused by tidal fluctuations may have muted diel patterns.

Sub-daily samples showed that adequate coverage of fish species was reached relatively quickly. For both sites and seasons, the sampling effort needed to optimize the fish species detections via eDNA reached an asymptote after three samples. Collecting more than three samples per day did not increase the probability of detecting more fish species' eDNA, and therefore may not be worth the additional collection and analysis costs. Autonomous platforms have a limited number of samples that can be collected and stored before needing replacement. For example, the Dartmouth Ocean Technologies eDNA Sampler can collect nine samples and the Woods Hole Oceanographic Large Volume eDNA sampler can collect 12 samples (Table 1 in Hendricks et al. 2023). At our study sites, three samples per day were sufficient to describe detectable fish species while minimizing cost and human time and effort.

We found that spreading out eDNA samples across days rather than within days is more critical to minimizing misrepresentation of communities. While the probability of detecting eDNA from all fish species was consistent on most days, there were five days that clearly had lower or higher detection probabilities, and thus observed species diversity, compared to the average day. In fact, at the Palisades Reservoir streamgage, there were two consecutive days when eDNA samples showed large swings in detection probability and diversity relative to the average. Day 208 resulted in below average diversity followed by day 209 when replicates averaged higher than expected detection probabilities. Similar to Hallam et al. (2023), who used shorter- and longer-term temporal sampling schemes to describe fish communities detected by eDNA, none of the environmental variables considered here explained this variation. Therefore, until more informative covariates are identified, we are unable to identify the mechanism behind these fluctuations and cannot predict which eDNA sampling days will accurately represent the fish communities. These sampling day effects indicate that eDNA samples should not be collected on a single day, or even on consecutive days, but rather be temporally dispersed. Spreading out eDNA samples across days will reduce the likelihood of misrepresenting fish diversity and may even improve detection of new biological threats (Sepulveda et al. 2021). Simulations of targeted eDNA detections showed similar results: dispersing eDNA sampling across days may be more effective at detecting new invasions whose DNA is likely to be at low abundances compared to increasing the number of samples collected within a day (Sepulveda et al. 2021). Likewise, Hallam

et al. (2023) provided similar guidance where they suggested eDNA biomonitoring should be spread across seasons rather than at a discrete time of year.

In contrast to other research assessing eDNA detections at a longer time scale (e.g., Hallam et al. 2023 and Searcy et al. 2022), we did not detect a difference of seasonal effect at either site, even when we account for the different primers used. In combination with the lack of statistical difference in eDNA detection probabilities for each species between seasons, these results indicate that the probability of detecting these fish species at the USGS streamgage site did not change between July and September even though their activity, relative abundance, and distribution in the reservoirs likely did. This may be a function of hydrodynamics and site features (i.e., reservoirs and dams). Water temperature and discharge were consistent during July and September relative to the rest of the year and did not appear to influence our results.

When using eDNA sampling to detect species of interest, it is important to consider which specific supplies (e.g., filter type and pore size) are needed to optimize the likelihood of detection. We found that the effect of pore size was minimal relative to other factors such as primer choice and sampling frequency. Previous research has shown mixed results where filter pore size for different filter types (i.e., MCE, Cellulose Nitrate, and Glass Fiber) may (Deiner et al. 2018) or may not (Bessey et al. 2020) have an impact on fish eDNA metabarcoding results. Similar to Bessey et al. (2020), we found no evidence of a difference in the probability of detection of each species due to filter pore size; however, we would likely only be able to detect large effects because there were only six paired sets of samples using each filter sizes of 0.45, 1.2, and 5.00 µm. These results held when we accounted for the differences in water volume filtered. Still, it is important to note that eDNA sampling is not a “one size fits all” approach and these results may not be applicable to all systems (e.g., Kumar et al. 2022). Before designing an eDNA biomonitoring program, site characteristics (e.g., turbidity) should be considered to optimize methods and supplies.

Conclusions

Metabarcoding of eDNA samples is becoming more commonly used and applied toward species detections and may be useful for future management (Hallam et al. 2023). Results from empirical studies like ours may help inform best practices for biomonitoring at USGS streamgages. In this study, we demonstrated the effectiveness of leveraging existing USGS streamgages for characterizing and monitoring fish diversity downstream of reservoirs. Reservoirs provide important services such as hydropower, flood control, and recreation, but can also be hotspots of invasive species introductions (Havel et al. 2005; Johnson et al. 2008). With more than 8,200 USGS streamgages located throughout the United States (Eberts et al. 2019), the incorporation of eDNA metabarcoding to characterize aquatic fauna could be a powerful way to monitor biodiversity. Results from this study may be useful in informing future sampling design for biomonitoring at USGS streamgages.

Finally, we demonstrate the usefulness of autonomous eDNA samplers to describe fish communities at USGS streamgages downstream of reservoirs. We showed that, of the factors we considered, primer choice and sampling day had the greatest impact on species eDNA detection probabilities. Autonomous platforms,

like the ESP, are particularly useful in that they can collect many eDNA samples. One current limitation of many autonomous platforms is their inability to collect multiple replicates at the same time. Additional replicates collected in parallel, either autonomously or manually, could account for the impact of time of day on species detections. When autonomous platforms are used in biodiversity monitoring programs and budgets are limited, spreading out samples over the course of several days rather than taking several samples in a short period of time may be optimal.

Acknowledgments

We would like to thank Dr. Donald Zaroban (College of Idaho), Jesse McCane (Idaho Fish and Game), and Matthew Campbell (Idaho Fish and Game) for providing tissue samples for mitochondrial sequencing. Additionally, we would like to thank Dr. Masayuki Ushio, Dr. Nathan Griffiths, and one anonymous reviewer for their thoughtful feedback which helped improve the manuscript.

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Funding

This project is one among a set of coordinated projects funded in part by the Bipartisan Infrastructure Law through the U.S. Department of the Interior to advance a nationally coordinated Early Detection and Rapid Response Framework. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. The authors have declared that no competing interests exist.

Author contributions

Conceptualization: DNJ, EB, PH, AJS. Data curation: PH, AJS, DNJ, BCA, RT, RR. Formal analysis: BCA, DNJ, RR. Funding acquisition: AJS, EB. Investigation: PH, AJS, SJ, EB, KMY, DNJ, JMB. Methodology: BCA, DNJ, PH, AJS. Project administration: AJS, PH, JRC, DNJ. Resources: EB, RT, JRC, AJS, JMB, KMY, SJ, RR. Software: JMB, RR, SJ, KMY, RT. Supervision: EB, AJS. Validation: PH, BCA, DNJ. Visualization: DNJ, BCA. Writing - original draft: DNJ, AJS. Writing - review and editing: AJS, EB, DNJ, BCA, JMB, PH, KMY, SJ, RT, JRC, RR.

Author ORCIDs

Devin N. Jones https://orcid.org/0000-0001-9215-2930
Ben C. Augustine https://orcid.org/0000-0001-6935-6361
Patrick Hutchins https://orcid.org/0000-0001-5232-0821
James M. Birch https://orcid.org/0000-0002-6080-8955
Kevan M. Yamahara https://orcid.org/0000-0003-3344-0283
Rodney Richardson https://orcid.org/0000-0002-4443-1705
James R. Campbell https://orcid.org/0000-0002-2760-3149
Elliott Barnhart https://orcid.org/0000-0002-8788-8393
Adam J. Sepulveda https://orcid.org/0000-0001-7621-7028

Data availability

Data and code are available at https://doi.org/10.5066/P1XKA9NV (Jones 2024). Metabarcoding sequence data are available at https://www.ncbi.nlm.nih.gov/sra/PRJNA1126731 (BioProject PRJNA1126731) under accessions SAMN41981679–SAMN41981966 (Jones et al. 2024).

References

Aichi Biodiversity Targets (2020) Convention on Biological Diversity. https://www.cbd.int/sp/targets/

Bálint M, Pfenninger M, Grossart H-P, Taberlet P, Vellend M, Leibold MA, Englund G, Bowler D (2018) Environmental DNA Time Series in Ecology. Trends in Ecology & Evolution 33(12): 945–957. https://doi.org/10.1016/j.tree.2018.09.003

Barnes MA, Turner CR (2016) The ecology of environmental DNA and implications for conservation genetics. Conservation Genetics 17(1): 1–17. https://doi.org/10.1007/s10592-015-0775-4

Beentjes KK, Speksnijder AGCL, Schilthuizen M, Hoogeveen M, Van Der Hoorn BB (2019) The effects of spatial and temporal replicate sampling on eDNA metabarcoding. PeerJ 7: e7335. https://doi.org/10.7717/peerj.7335

Bessey C, Jarman SN, Berry O, Olsen YS, Bunce M, Simpson T, Power M, McLaughlin J, Edgar GJ, Keesing J (2020) Maximizing fish detection with eDNA metabarcoding. Environmental DNA 2(4): 493–504. https://doi.org/10.1002/edn3.74

Betancourt M, Girolami M (2015) Hamiltonian monte carlo for hierarchical models. Current trends in Bayesian methodology with applications 79: 2–4. https://doi.org/10.1201/b18502-5

Bista I, Carvalho GR, Walsh K, Seymour M, Hajibabaei M, Lallias D, Christmas M, Creer S (2017) Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. Nature Communications 8(1): 14087. https://doi.org/10.1038/ncomms14087

Bylemans J, Gleeson DM, Hardy CM, Furlan E (2018) Toward an ecoregion scale evaluation of eDNA metabarcoding primers: A case study for the freshwater fish biodiversity of the Murray–Darling Basin (Australia). Ecology and Evolution 8(17): 8697–8712. https://doi.org/10.1002/ece3.4387

Crone MK, Boyle NK, Bresnahan ST, Biddinger DJ, Richardson RT, Grozinger CM (2023) More than mesolectic: Characterizing the nutritional niche of *Osmia cornifrons*. Ecology and Evolution 13(10): e10640. https://doi.org/10.1002/ece3.10640

de Valpine P, Turek D, Paciorek CJ, Anderson-Bergman C, Lang DT, Bodik R (2017) Programming with models: Writing statistical algorithms for general model structures with nimble. Journal of Computational and Graphical Statistics 26(2): 403–413. https://doi.org/10.1080/10618600.2016.1172487

Deiner K, Lopez J, Bourne S, Holman L, Seymour M, Grey EK, Lacoursière-Roussel A, Li Y, Renshaw MA, Pfrender ME, Rius M, Bernatchez L, Lodge DM (2018) Optimising the detection of marine taxonomic richness using environmental DNA metabarcoding: The effects of filter material, pore size and extraction method. Metabarcoding and Metagenomics 2: e28963. https://doi.org/10.3897/mbmg.2.28963

Díaz S, Settele J, Brondízio E, Ngo HT, Guèze M, Agard J, Arneth A, Balvanera P, Brauman KA, Butchart SHM, Chan KMA, Garibaldi LA, Ichii K, Liu J, Subramanian SM, Midgley GF, Miloslavich P, Molnár Z, Obura D, Pfaff A, Polasky S, Purvis A, Razzaque J, Reyers B, Chowdhury RR, Shin YJ, Visseren-Hamakers IJ, Willis KJ, Zayas CN (2019) IPBES

(2019) Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Bonn, Germany.

Dorazio RM, Royle JA, Söderström B, Glimskär A (2006) EStimating species richness and accumulation by modeling species occurrence and detectability. Ecology 87(4): 842–854. https://doi.org/10.1890/0012-9658(2006)87[842:ESRAAB]2.0.CO;2

Dziedzic E, Sidlauskas B, Cronn R, Anthony J, Cornwell T, Friesen TA, Konstantinidis P, Penaluna BE, Stein S, Levi T (2023) Creating, curating and evaluating a mitogenomic reference database to improve regional species identification using environmental DNA. Molecular Ecology Resources 23(8): 1880–1904. https://doi.org/10.1111/1755-0998.13855

Eberts SM, Woodside MD, Landers MN, Wagner CR (2019) Monitoring the pulse of our Nation's rivers and streams—The U.S. Geological Survey streamgaging network. Reston, VA. https://doi.org/10.3133/fs20183081

Eddy SR (2011) Accelerated Profile HMM Searches. PLoS Computational Biology 7(10): e1002195. https://doi.org/10.1371/journal.pcbi.1002195

Gelman A, Hill J (2006) Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, 648 pp. https://doi.org/10.1017/CBO9780511790942

Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. Statistical Science 7(4): 457–472. https://doi.org/10.1214/ss/1177011136

Gunn JM (1995) Spawning behavior of lake trout: Effects on colonization ability. Journal of Great Lakes Research 21: 323–329. https://doi.org/10.1016/S0380-1330(95)71106-6

Hajibabaei M, Smith MA, Janzen DH, Rodriguez JJ, Whitfield JB, Hebert PDN (2006) A minimalist barcode can identify a specimen whose DNA is degraded. Molecular Ecology Notes 6(4): 959–964. https://doi.org/10.1111/j.1471-8286.2006.01470.x

Hajibabaei M, Porter TM, Wright M, Rudar J (2019) COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems. PLoS ONE 14(9): e0220953. https://doi.org/10.1371/journal.pone.0220953

Hallam J, Clare EL, Jones JI, Day JJ (2023) High frequency environmental DNA metabarcoding provides rapid and effective monitoring of fish community dynamics. Environmental DNA edn3.486. https://doi.org/10.1002/edn3.486

Handley LL, Read DS, Winfield IJ, Kimbell H, Johnson H, Li J, Hahn C, Blackman R, Wilcox R, Donnelly R, Szitenberg A, Hänfling B (2019) Temporal and spatial variation in distribution of fish environmental DNA in England's largest lake. Environmental DNA 1(1): 26–39. https://doi.org/10.1002/edn3.5

Havel JE, Lee CE, Vander Zanden MJ (2005) Do reservoirs facilitate invasions into landscapes? Bioscience 55(6): 518. https://doi.org/10.1641/0006-3568(2005)055[0518:DRFIIL]2.0.CO;2

Hayami K, Sakata MK, Inagawa T, Okitsu J, Katano I, Doi H, Nakai K, Ichiyanagi H, Gotoh RO, Miya M, Sato H, Yamanaka H, Minamoto T (2020) Effects of sampling seasons and locations on fish environmental DNA metabarcoding in dam reservoirs. Ecology and Evolution 10(12): 5354–5367. https://doi.org/10.1002/ece3.6279

Henderson R, Kershner JL, Toline CA (2000) Timing and location of spawning by nonnative wild rainbow trout and native cutthroat trout in the South Fork Snake River, Idaho, with implications for hybridization. North American Journal of Fisheries Management 20(3): 584–596. https://doi.org/10.1577/1548-8675(2000)020<0584:TALOSB>2.3.CO;2

Hendricks A, Mackie CM, Luy E, Sonnichsen C, Smith J, Grundke I, Tavasoli M, Furlong A, Beiko RG, LaRoche J, Sieben V (2023) Compact and automated eDNA sampler for in

situ monitoring of marine environments. Scientific Reports 13(1): 5210. https://doi.org/10.1038/s41598-023-32310-3

Hervé A, Domaizon I, Baudoin JM, Dejean T, Gibert P, Jean P, Peroux T, Raymond J-C, Valentini A, Vautier M, Logez M (2022) Spatio-temporal variability of eDNA signal and its implication for fish monitoring in lakes. PLoS ONE 17(8): e0272660. https://doi.org/10.1371/journal.pone.0272660

Jensen MR, Sigsgaard EE, Ávila MDP, Agersnap S, Brenner-Larsen W, Sengupta ME, Xing Y, Møller PR, Thomsen PF (2022) Short-term temporal variation of coastal marine eDNA. Environmental DNA 4(4): 747–762. https://doi.org/10.1002/edn3.285

Jo T, Murakami H, Masuda R, Minamoto T (2020) Selective collection of long fragments of environmental DNA using larger pore size filter. The Science of the Total Environment 735: 139462. https://doi.org/10.1016/j.scitotenv.2020.139462

Johnson PT, Olden JD, Vander Zande MJ (2008) Dam invaders: Impoundments facilitate biological invasions into freshwaters. Frontiers in Ecology and the Environment 6(7): 357–363. https://doi.org/10.1890/070156

Jones DN (2024) Metabarcoding of environmental DNA samples collected below Jackson Lake Dam (Wyoming) and Palisades Reservoir (Idaho) in 2021-2022.U.S. Geological Survey data release. https://doi.org/10.5066/P1XKA9NV

Jones DN, Augustine BC, Hutchins P, Birch JM, Yamahara KM, Jensen S, Richardson R, Trott R, Campbell JR, Barnhart E, Sepulveda AJ (2024) Autonomous samplers and environmental DNA metabarcoding: sampling day and primer choice have greatest impact on fish detection probabilities. Accessions SAMN41981679–SAMN41981966. GenBank Submission (BioProject PRJNA1126731). https://www.ncbi.nlm.nih.gov/sra/PRJNA1126731

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30(14): 3059–3066. https://doi.org/10.1093/nar/gkf436

Keck F, Couton M, Altermatt F (2023) Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. Molecular Ecology Resources 23(4): 742–755. https://doi.org/10.1111/1755-0998.13746

Kumar G, Farrell E, Reaume AM, Eble JA, Gaither MR (2022) One size does not fit all: Tuning eDNA protocols for high- and low-turbidity water sampling. Environmental DNA 4(1): 167–180. https://doi.org/10.1002/edn3.235

Mariani S, Fernandez C, Baillie C, Magalon H, Jaquemet S (2021) Shark and ray diversity, abundance and temporal variation around an Indian Ocean Island, inferred by eDNA metabarcoding. Conservation Science and Practice 3(6): e407. https://doi.org/10.1111/csp2.407

Mathieu C, Hermans SM, Lear G, Buckley TR, Lee KC, Buckley HL (2020) A Systematic Review of Sources of Variability and Uncertainty in eDNA Data for Environmental Monitoring. Frontiers in Ecology and Evolution 8: 135. https://doi.org/10.3389/fevo.2020.00135

McKnight DT, Huerlimann R, Bower DS, Schwarzkopf L, Alford RA, Zenger KR (2019) microDecon: A highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies. Environmental DNA 1(1): 14–25. https://doi.org/10.1002/edn3.11

Meusnier I, Singer GA, Landry J-F, Hickey DA, Hebert PD, Hajibabaei M (2008) A universal DNA mini-barcode for biodiversity analysis. BMC Genomics 9(1): 214. https://doi.org/10.1186/1471-2164-9-214

Miya M, Sato Y, Fukunaga T, Sado T, Poulsen JY, Sato K, Minamoto T, Yamamoto S, Yamanaka H, Araki H, Kondoh M, Iwasaki W (2015) MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: Detection of more than 230 subtropical marine species. Royal Society Open Science 2(7): 150088. https://doi.org/10.1098/rsos.150088

Miya M, Gotoh RO, Sado T (2020) MiFish metabarcoding: A high-throughput approach for simultaneous detection of multiple fish species from environmental DNA and other samples. Fisheries Science 86(6): 939–970. https://doi.org/10.1007/s12562-020-01461-x

Moussy C, Burfield IJ, Stephenson PJ, Newton AFE, Butchart SHM, Sutherland WJ, Gregory RD, McRae L, Bubb P, Roesler I, Ursino C, Wu Y, Retief EF, Udin JS, Urazaliyev R, Sánchez-Clavijo LM, Lartey E, Donald PF (2022) A quantitative global review of species population monitoring. Conservation Biology 36(1): 1–14. https://doi.org/10.1111/cobi.13721

Nowak GM, Quinn TP (2002) Diel and seasonal patterns of horizontal and vertical movements of telemetered cutthroat trout in Lake Washington, Washington. Transactions of the American Fisheries Society 131(3): 452–462. https://doi.org/10.1577/1548-8659(2002)131<0452:DASPOH>2.0.CO;2

Pawlowski J, Bonin A, Boyer F, Cordier T, Taberlet P (2021) Environmental DNA for biomonitoring. Molecular Ecology 30(13): 2931–2936. https://doi.org/10.1111/mec.16023

Pont D, Rocle M, Valentini A, Civade R, Jean P, Maire A, Roset N, Schabuss M, Zornig H, Dejean T (2018) Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. Scientific Reports 8(1): 10361. https://doi.org/10.1038/s41598-018-28424-8

Preston C, Yamahara K, Pargett D, Weinstock C, Birch J, Roman B, Jensen S, Connon B, Jenkins R, Ryan J, Scholin C (2023) Autonomous eDNA collection using an uncrewed surface vessel over a 4200-km transect of the eastern Pacific Ocean. Environmental DNA edn3.468. https://doi.org/10.1002/edn3.468

Reebs SG, Boudreau L, Hardie P, Cunjak RA (1995) Diel activity patterns of lake chubs and other fishes in a temperate stream. Canadian Journal of Zoology 73(7): 1221–1227. https://doi.org/10.1139/z95-146

Richardson RT, Curtis HR, Matcham EG, Lin C-H, Suresh S, Sponsler DB, Hearon LE, Johnson RM (2019) Quantitative multi-locus metabarcoding and waggle dance interpretation reveal honey bee spring foraging patterns in Midwest agroecosystems. Molecular Ecology 28(3): 686–697. https://doi.org/10.1111/mec.14975

Richardson RT, DB Sponsler, McMinn-Sauder H, RM Johnson (2020) MetaCurator: A hidden Markov model-based toolkit for extracting and curating sequences from taxonomically-informative genetic markers. Methods in Ecology and Evolution 11: 181–186. https://doi.org/10.1111/2041-210X.13314

Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: A versatile open source tool for metagenomics. PeerJ 4: e2584. https://doi.org/10.7717/peerj.2584

Ruppert KM, Kline RJ, Rahman MS (2019) Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. Global Ecology and Conservation 17: e00547. https://doi.org/10.1016/j.gecco.2019.e00547

Sales NG, Wangensteen OS, Carvalho DC, Deiner K, Præbel K, Coscia I, McDevitt AD, Mariana S (2021) Space-time dynamics in monitoring neotropical fish communities

using eDNA metabarcoding. The Science of the Total Environment 754: 142096. https://doi.org/10.1016/j.scitotenv.2020.142096

Schenekar T, Schletterer M, Lecaudey LA, Weiss SJ (2020) Reference databases, primer choice, and assay sensitivity for environmental metabarcoding: Lessons learnt from a re-evaluation of an eDNA fish assessment in the Volga headwaters. River Research and Applications 36(7): 1004–1013. https://doi.org/10.1002/rra.3610

Scholin C, Birch J, Jensen S, Marin R III, Massion E, Pargett D, Preston C, Roman B, Ussler W III (2017) The quest to develop ecogenomic sensors: A 25-year history of the environmental sample processor (ESP) as a case study. Oceanography (Washington, D.C.) 30(4): 100–113. https://doi.org/10.5670/oceanog.2017.427

Searcy RT, Boehm AB, Weinstock C, Preston CM, Jensen S, Roman B, Birch JM, Scholin CA, Van Houtan KS, Kiernan JD, Yamahara KM (2022) High-frequency and long-term observations of eDNA from imperiled salmonids in a coastal stream: Temporal dynamics, relationships with environmental factors, and comparisons with conventional observations. Environmental DNA 4(4): 776–789. https://doi.org/10.1002/edn3.293

Sepulveda AJ, Schmidt C, Amberg J, Hutchins P, Stratton C, Mebane C, Laramie MB, Pilliod DS (2019) Adding invasive species biosurveillance to the U.S. Geological Survey streamgage network. Ecosphere 10(8): e02843. https://doi.org/10.1002/ecs2.2843

Sepulveda AJ, Birch JM, Barnhart EP, Merkes CM, Yamahara KM, Marin R III, Kinsey SM, Wright PR, Schmidt C (2020) Robotic environmental DNA bio-surveillance of freshwater health. Scientific Reports 10(1): 14389. https://doi.org/10.1038/s41598-020-71304-3

Sepulveda AJ, Hoegh A, Gage JA, Caldwell Eldridge SL, Birch JM, Stratton C, Hutchins PR, Barnhart EP (2021) Integrating environmental DNA results with diverse data sets to improve biosurveillance of river health. Frontiers in Ecology and Evolution 9: 620715. https://doi.org/10.3389/fevo.2021.620715

Tillotson MD, Kelly RP, Duda JJ, Hoy M, Kralj J, Quinn TP (2018) Concentrations of environmental DNA (eDNA) reflect spawning salmon abundance at fine spatial and temporal scales. Biological Conservation 220: 1–11. https://doi.org/10.1016/j.biocon.2018.01.030

U.S. Geological Survey (2023) National Water Information System data available on the World Wide Web (USGS Water Data for the Nation) [accessed May 30, 2023 http://waterdata.usgs.gov/nwis/]. https://doi.org/10.5066/F7P55KJN

Westfall KM, Singer GAC, Kaushal M, Gilmore SR, Fahner N, Hajibabaei M, Abbott CL (2023) NAMERS: A purpose-built reference DNA sequence database to support applied eDNA metabarcoding. Genetics. https://doi.org/10.1101/2023.10.06.561210

Zhang S, Zhao J, Yao M (2020) A comprehensive and comparative evaluation of primers for metabarcoding eDNA from fish. Methods in Ecology and Evolution 11(12): 1609–1625. https://doi.org/10.1111/2041-210X.13485

Supplementary material 1

Additional detail for methodology as well as additional figures

Authors: Devin N. Jones, Ben C. Augustine, Patrick Hutchins, James M. Birch, Kevan M. Yamahara, Scott Jensen, Rodney Richardson, Regina Trott, James R. Campbell, Elliott Barnhart, Adam J. Sepulveda

Data type: docx

Link: https://doi.org/10.3897/mbmg.8.122375.suppl1

Supplementary material 2

Sequence data for mock communities used in metabarcoding

Authors: Devin N. Jones, Ben C. Augustine, Patrick Hutchins, James M. Birch, Kevan M. Yamahara, Scott Jensen, Rodney Richardson, Regina Trott, James R. Campbell, Elliott Barnhart, Adam J. Sepulveda

Data type: fasta

Link: https://doi.org/10.3897/mbmg.8.122375.suppl2

Supplementary material 3

Summaries of reads removed during the decontamination process

Authors: Devin N. Jones, Ben C. Augustine, Patrick Hutchins, James M. Birch, Kevan M. Yamahara, Scott Jensen, Rodney Richardson, Regina Trott, James R. Campbell, Elliott Barnhart, Adam J. Sepulveda

Data type: xlsx

Link: https://doi.org/10.3897/mbmg.8.122375.suppl3